

学習における教師による誤り修正の効果

西 正明

降矢 順治

A Effect of Correcting Error in Learning with a Teacher

Masaaki Nishi Jyunji Furuya

(1990.12.25 受理)

キーワード：C A I , 学習, ニューラルネットワーク

1. 緒 言

現在、脳の構造を模したニューラルネットワークの研究が世界的に盛んに行われている。本論文では前報⁽¹⁾に引き続き、C A I を用いて学習を効率的に行うための基礎的研究として、ニューラルネットワークを用いた学習モデルのシミュレーションを行って学習時の挙動を検討した結果を述べる。学習モデルとしては階層型ニューラルネットワーク、学習アルゴリズムとしてバックプロパゲーションを用いてその特性を評価した。学習の対象として0から9までの数字を用い、特に教師による学習の誤り修正の効果について評価検討した。

2. 学習モデルとその評価

ニューラルネットワークによる学習モデルの概要を述べてから、この学習モデルの記憶特性の評価結果を述べる。さらに、この学習モデルから得られた学習に関する知見及び学習における教師の効果を述べる。

2. 1 学習モデルの概要

ニューラルネットワークにはその結合形式から相互結合型ネットワークと階層型ネットワークがある。相互結合型ネットワークは全ての神経細胞が相互に接続した構造のものであり、階層型ネットワークは細胞が階層状に規則正しく並んだもので隣合う層間でのみ結合し下位の細胞層から上位の細胞層に一方向にだけ情報が伝達される。階層型ネットワークはフィードバック機能を付加した帰還型ネッ

トワークにすると相互結合型ネットワークと同じ機能を持つことになるので、本論文では階層型ネットワークと帰還型ネットワークを取り扱う。

表1 入力パタンと期待される出力パタン

I	入力パタン番号	0 1 2 3 4 5 6 7 8 9
	期待される出力パタン番号	0 1 2 3 4 5 6 7 8 9
II	入力パタン番号	10 11 12 13 14 15 16 17
	期待される出力パタン番号	1 2 3 3 5 8 9 9 0
III	入力パタン番号	18 19 20 21 22 23 24
	期待される出力パタン番号	1 2 2 3 5 8 9

注 期待される出力パタンはパタン間の相関(内積値)から判断した

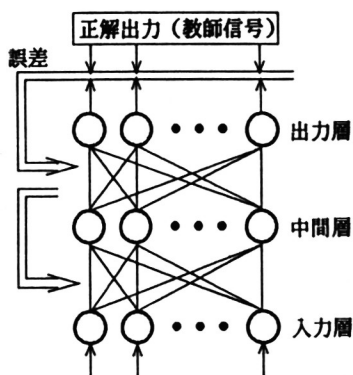


図1 バックプロパゲーション型学習

学習方法はバックプロパゲーションと呼ばれるアルゴリズムを用いる。図1に示すように、入力から出力を求め教師信号との差(誤差)をとり出力層から入力層へとネットワークを逆に辿り誤差が少なくなるようにネットワークの結合の重み値を修正する。教師信号は階層型ネットワークでは別々に与えられる(教師付き学習と呼ぶ)が、帰還型ネットワークでは入力と同じものを用いる(教師無し学習と呼ぶ)。この操作を多くのパタンについて繰り返し行う。

いま m 層のネットワークを考え、 k 層の第 j 番目の細胞への入力の総和を $i(j^k)$ 、出力を $o(j^k)$ とし、 $k-1$ 層の第 i 番目の細胞と k 層の第 j 番目の細胞との結合の重みを $w(i^{k-1}j^k)$ とする。各細胞の入出力関係を与える関数を f とすると、以上の変数間には次の関係がある。

$$\left. \begin{aligned} o(j^k) &= f(i(j^k)) \\ i(j^k) &= \sum_i w(i^{k-1}j^k) \cdot o(i^{k-1}) \end{aligned} \right\} \quad (1)$$

ある入力パタン x と教師パタン y が与えられたときの誤差 r は二乗誤差で示すと次式になる。

$$r = \sum_j (o(j^m, w, x) - y(j))^2 \quad (2)$$

ここで w はネットワークの重みを全てまとめて示すものとする。 w の修正量を求めるには、この r の w についての微分値を求めれば良い。

$$\begin{aligned} \Delta w(i^{k-1}j^k) &= - \frac{\partial r}{\partial w(i^{k-1}j^k)} \\ &= - \frac{\partial r}{\partial i(j^k)} \cdot \frac{\partial i(j^k)}{\partial w(i^{k-1}j^k)} \\ &= - \frac{\partial r}{\partial i(j^k)} \cdot o(i^{k-1}) \end{aligned} \quad (3)$$

$k \neq m$ のとき

$$\begin{aligned} \frac{\partial r}{\partial i(j^k)} &= \left(\sum_l \frac{\partial r}{\partial i(l^{k+1})} \cdot \frac{\partial i(l^{k+1})}{\partial o(j^k)} \right) \cdot \frac{\partial o(j^k)}{\partial i(j^k)} \\ &= \left(\sum_l \frac{\partial r}{\partial i(l^{k+1})} \cdot w(j^k l^{k+1}) \right) \cdot f'(i(j^k)) \end{aligned} \quad (4)$$

なので、 $\frac{\partial r}{\partial i(j^k)} = d(j^k)$ とおくと、結合の重みの修正量 $\Delta w(i^{k-1} j^k)$ は次式を満たすものとなる。

$$\begin{aligned} \Delta w(i^{k-1} j^k) &= -d(j^k) \cdot o(i^{k-1}) \\ d(j^m) &= (o(j^m) - y(j)) \cdot f'(i(j^m)) \\ d(j^k) &= \left(\sum_l w(j^k l^{k+1}) \cdot d(l^{k+1}) \right) \cdot f'(i(j^k)) \end{aligned} \quad (5)$$

関数 f は容易に微分できることが必要で通常シグモイド関数と呼ばれる次の式で定義される。

$$f(x) = \frac{1}{1 + \exp(-x + \theta)} \quad (6)$$

ここで θ はしきい値である。 $f'(x) = f(x) \cdot (1 - f(x))$ なので、

$$f'(i(j^k)) = o(j^k) \cdot (1 - o(j^k)) \text{ となる。}$$

また、学習が振動することなく最少の学習回数で最大の学習効果が得られるように、さらに以下の対策を施した。

① 学習の振動を少なくするために前回の修正量と今回の修正量に重み付けをした和を最終的な修正量とする。

$$\Delta w = a \Delta w_{old} + b \Delta w_{new} \quad (7)$$

ここで a は過去の重み、 b は学習速度である。

② 式(6)におけるしきい値 θ を学習時に修正する。修正量は常に 1 を出力する細胞との結合の重みとみなして式(5)で計算される。また、①と同様に次式を用いて修正量を抑制する。

$$\Delta \theta = c \Delta \theta_{new} \quad (8)$$

ここで c は学習速度である。

③ 学習データを周期的に繰り返し用いると、学習が同じことの繰り返しになって収束しなくなる場合がある。これを避けるために次式のように結合の重み値 w 、細胞の出力値 o にランダムなノイズを加える。

$$\left. \begin{aligned} w &= (\text{本来の } w) \cdot (1 + \alpha F) \\ o &= (\text{本来の } o) \cdot (1 + \alpha F) \end{aligned} \right\} \quad (9)$$

ここで F は正規分布形の乱数であり、 α はノイズ強度である。

④ 学習データが周期的に繰り返し現れるのを避けるために、出力層の誤差が指定した打ち切り誤差 e 以下になるまで同じ学習データを繰り返し用いて学習する。

2. 2 学習モデルの評価

評価に用いた入力データは、前報⁽¹⁾と同様図2に示す0～9を示すデータ（パタン番号0～9の10種）及びそれらにノイズを混入させたデータ（パタン番号10～24の15種）の合計25種のデータである。これらのパタンデータを入力したときパタン番号0～9との相関からみて期待される出力パタンを表1に示す。表1においてグループIに含まれる入力パタン（0～9）は後に述べるどの学習モデルでも必ず学習するパタン、グループIIに含まれる入力パタン（10～17）は教師付き学習でグループIと一緒に学習する学習パタンであり、グループIIIの入力パタン（18～24）は学習には使用せず評価にのみ使用するパタンである。

本論文では実際には、CRCから販売されているニューラルネットワークシミュレーションプログラムRHINEを利用して評価検討した⁽²⁾。前節2.1に述べた各種パラメータは、 $a=0.8$, $b=0.2$, $c=0.2$, $\alpha=0.1$, $e=0.2$ に設定し、誤差は出力層の細胞が n 個、 i 番目の細胞の出力が y_i , 正解が t_i としたとき次式で算出される二乗平均誤差とした。

$$\text{二乗平均誤差} = \left(\sum_i (y_i - t_i)^2 / n \right)^{1/2} \quad (10)$$

評価の手順としては、学習モデルの基本構成即ち層数、細胞数を決め、学習順番を決める。次に細胞、結合の重みにそれぞれノイズの付加をする／しない、しきい値学習をする／しないの計8通りの組み合わせについて学習の収束状況を比較評価する。教師付き学習についても同様に8通りの組み合わせについて学習の収束状況を調べ、教師無し学習との比較評価をする。最後に教師無し学習と教師付き学習とを混在させる場合についても検討する。従って評価すべき学習モデルは表2にまとめられる。

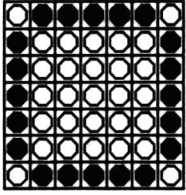
以下に比較評価した結果を述べる。

(1) 学習モデルの構成

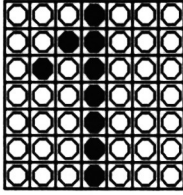
層数は3層または4層であればあらゆる問題に対応できるとされている⁽³⁾。本論文では評価を簡単にするため3層構造とした。細胞数は図形入力、図形出力なので入力層、出力層とも図2に示すパタンの格子数49と等しくした。中間層の細胞については9, 16, 25, 36, 49, 64の6種について学習の収束状況を比較した。結果を図3に示す。図3において縦軸の正答率とは出力層の各細胞の出力値（式(6)に従った0～1の値）を0.5以上で1, 0.5未満で0に置き換えたときに表示される出力パタンと表1に示した期待される出力パタンとを比較して、全評価パタンについて一致しているパタン数の割合を示すものとする。図3からどの場合も学習回数と共に正答率は向上するのがわかる。また、中間層の細胞数（図中記号S3n0は中間層の細胞数が n^2 であることを示す）は少なくとも多くても結果は悪くなり適切な数が存在することがわかる。ここで扱った問題に関しては、図中S350即ち中間層の細胞数は25が適切と判断した。

なお、学習回数については学習データの現れ方を観察していると200回～500回でようやく全学習データを一通り学習し終え、1000回程度でほぼ収束し、5000回以上では過学習の現象が起きる場合があった。

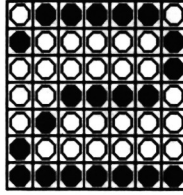
パターン番号 0



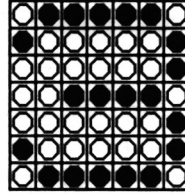
パターン番号 1



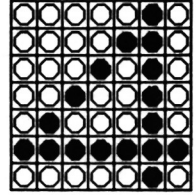
パターン番号 2



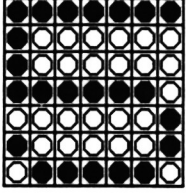
パターン番号 3



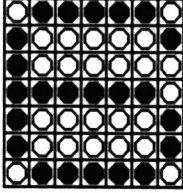
パターン番号 4



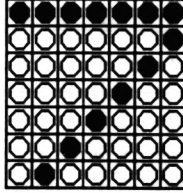
パターン番号 5



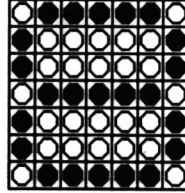
パターン番号 6



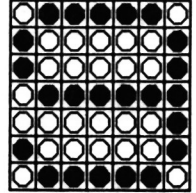
パターン番号 7



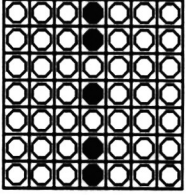
パターン番号 8



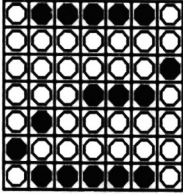
パターン番号 9



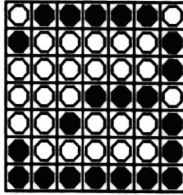
パターン番号 10



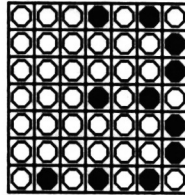
パターン番号 11



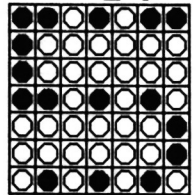
パターン番号 12



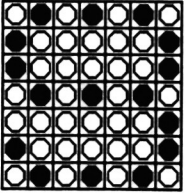
パターン番号 13



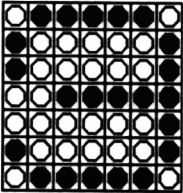
パターン番号 14



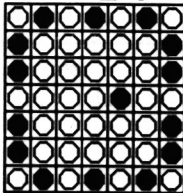
パターン番号 15



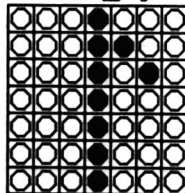
パターン番号 16



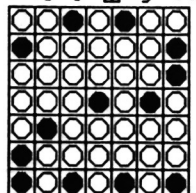
パターン番号 17



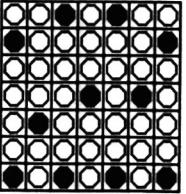
パターン番号 18



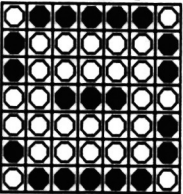
パターン番号 19



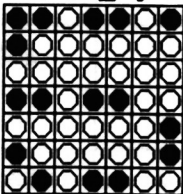
パターン番号 20



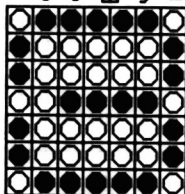
パターン番号 21



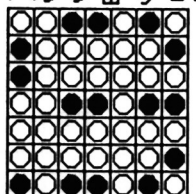
パターン番号 22



パターン番号 23



パターン番号 24



● : 1 , ○ : 0

図2 パタン番号とパタン

表2 評価学習モデル一覧

モデル名	中間層細胞数	学習順番	ノイズ付加	しきい値学習	教師付学習
S330	9	ランダム	しない	しない	しない
S340	16	〃	〃	〃	〃
S350	25	〃	〃	〃	〃
S360	36	〃	〃	〃	〃
S370	49	〃	〃	〃	〃
S380	64	〃	〃	〃	〃
S350C	25	イクリック	しない	しない	しない
S350S	〃	ランダムスキャン	〃	〃	〃
S351	25	ランダム	しない	する	しない
S352	〃	〃	細胞	しない	〃
S353	〃	〃	〃	する	〃
S354	〃	〃	重み値	しない	〃
S355	〃	〃	〃	する	〃
S356	〃	〃	細胞, 重み値	しない	〃
S357	〃	〃	〃	する	〃
T350	25	ランダム	しない	しない	する
T351	〃	〃	〃	する	〃
T352	〃	〃	細胞	しない	〃
T353	〃	〃	〃	する	〃
T354	〃	〃	重み値	しない	〃
T355	〃	〃	〃	する	〃
T356	〃	〃	細胞, 重み値	しない	〃
T357	〃	〃	〃	する	〃

(2) 学習順番

学習順番は、学習データを順番に周期的に繰り返すサイクリック、学習データの順番を周期毎にランダムに入れ換えるランダムスキャン及び学習データを全く不規則に組み換えるランダムの3方法について学習の収束状況を比較評価した。評価結果を図4に示す。S350はランダム、S350Cはサイクリック、S350Sはランダムスキャンであり、これらを比較するとサイクリック学習S350Cは正答率が低く100%には達しないのに対し、ランダム学習S350は最も正答率が高く学習順番としては適切であると判断できる。

(3) ノイズ付加, しきい値学習

(1), (2)の結果から、細胞数が49-25-49の3層構造のニューラルネットワークによる学習モデルを用いて学習データをランダムに与えて学習を行うことにして、ニューラルネットワークの各細胞の出力値、結合の重み値にランダムノイズを付加する効果、しきい値学習の効果を調べる。

帰還型ネットワークを用いた教師無し学習について正答率を評価した結果を図5である。これより学習モデルによって学習の収束性に差が生じ、正答率が100%に達していることからS350, S353, S355が良さそうであることがわかる。また、S354~S357は結合の重みにランダムノイズを付加しながら学習を進めるため処理量が多いのであるが、その割に正答率は向上していないことがわかる。また、しきい値学習を行ったことによる有意差は認められなかった。次に、階層型ネットワークを用いた教師付き学習について同様に評価した結果を図6に示す。図6ではノイズ付加、しきい値学習の有無に依らず正答率は100%に達しており、図5と比較すると教師付き学習が有効であることがわかる。

さらに、図5または、図6のいくつかの学習モデルについての各評価パタンの二乗平均誤差を図7.1, 図7.2に示す。図7.1および図7.2から教師付き学習(図注T***)は教師無し学習(図中S***)に比べて全評価パターンに渡って一様に二乗平均誤差が小さく教師付き学習の有効性が示された。

(4) 教師無し学習と教師付き学習との混在

しきい値学習あり、ノイズ付加なしの場合を例に学習回数が1000回を境に教師無し学習と教師付き学習を切り替えて両者を混在させた場合と全てを教師無し学習または教師付き学習を行う場合を比較した。図8にそれらの正答率と平均二乗平均誤差(二乗平均誤差を全評価パターンについてさらに平均した値)を示す。図8から全ての学習を教師無しで行う場合S351に比べて教師付きで行う場合T351は正答率、平均二乗平均誤差共に優れている。前半教師付き学習を行い後半教師無し学習を行う場合TS351はT351とほとんど差がなかった。前半教師無し後半教師付き学習を行う場合ST351は、後半教師付き学習に切り換えるとT351に正答率、平均二乗平均誤差共に接近することがわかる。これより部分的にでも教師付き学習を行うことは学習効率を上げるのに有効であることがわかる。

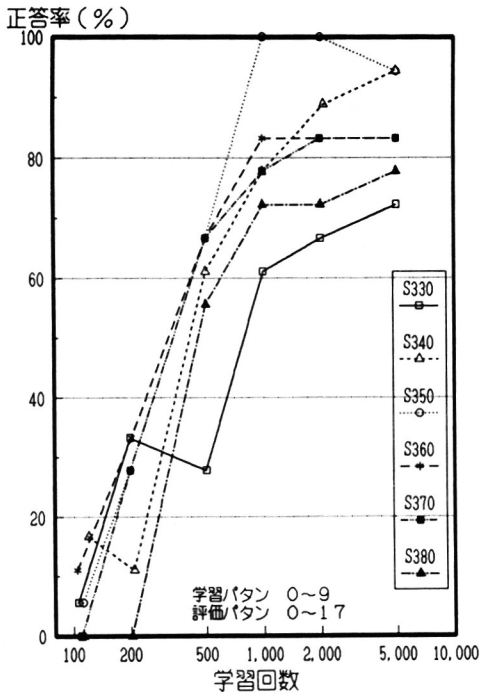


図3 中間層細胞数と正答率

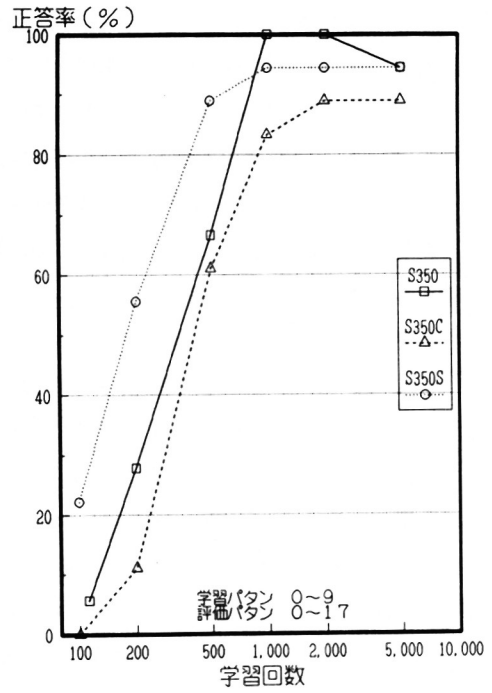


図4 学習順番と正答率

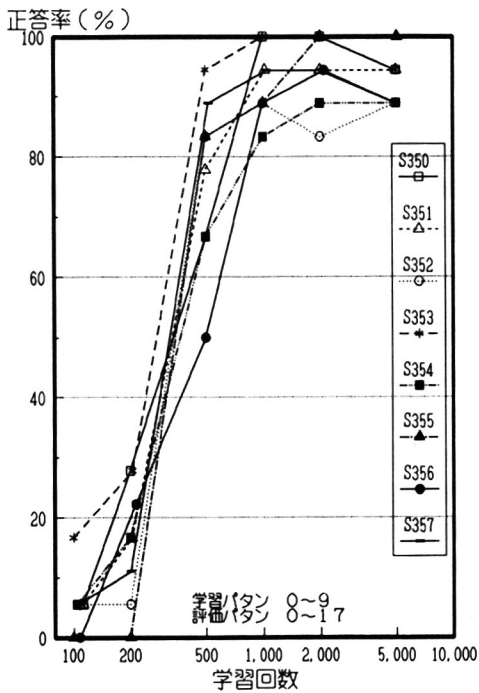


図5 ノイズ付加, しきい値学習と正答率

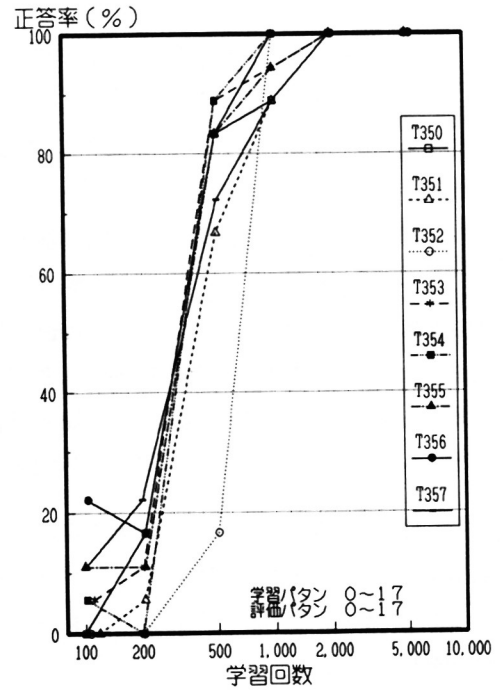


図6 教師付き学習における
ノイズ付加, しきい値学習と正答率

二乗平均誤差

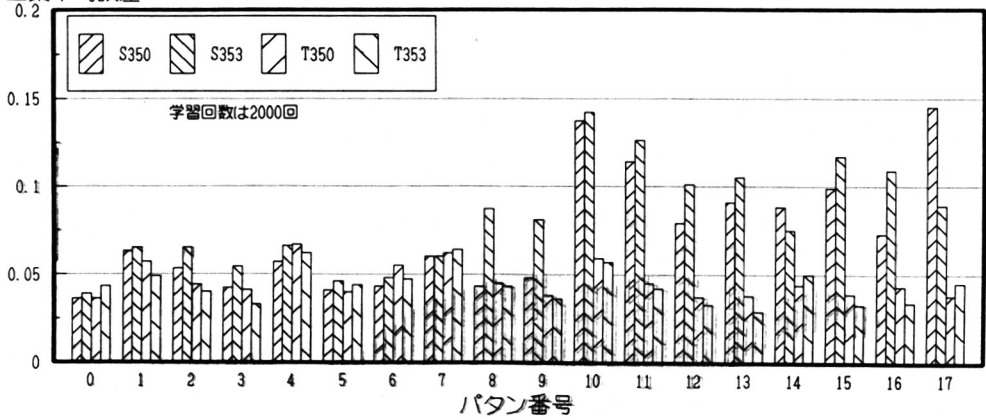


図7.1 パタン番号と二乗平均誤差 (X350, X353)

二乗平均誤差

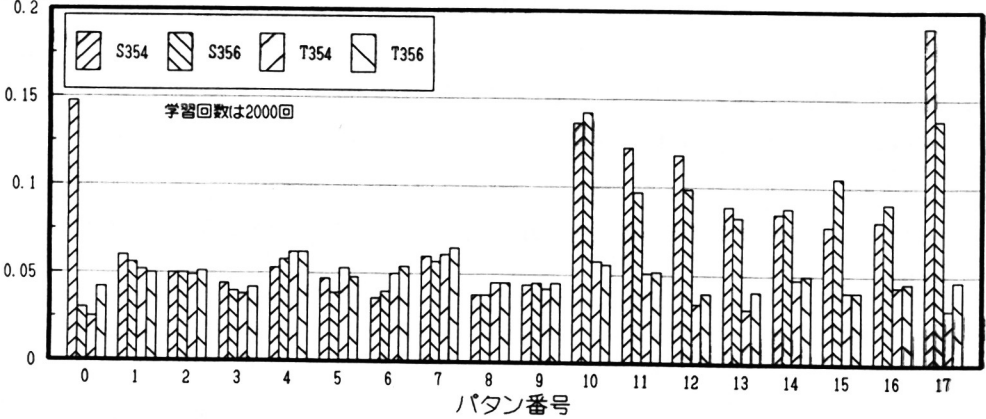


図7.2 パタン番号と二乗平均誤差 (X354, X356)

正答率 (%)

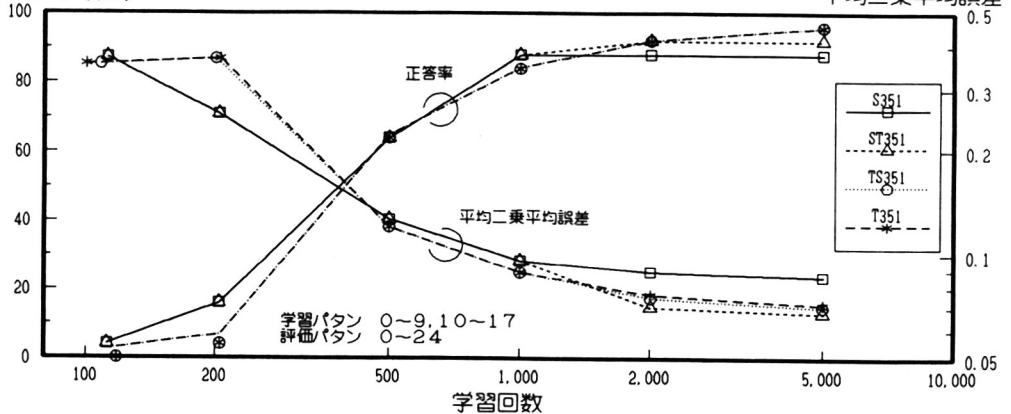


図8 教師無し学習と教師付き学習との混在

3. 結 言

脳の構造を模したニューラルネットワーク特に階層型ニューラルネットワークと帰還型ニューラルネットワークを用いて、バックプロパゲーションと呼ばれる学習アルゴリズムに従って0～9の数字を学習する際の学習特性を評価検討した。その結果、ニューラルネットワークとしては細胞数が多いほど学習がより完全に行われるとは限らず、適切な細胞数が存在することがわかった。学習順番としては、繰り返し学習するにしても常に同じ順番で学習するよりは、ランダムに入れ換えて学習する方が効果的であることがわかった。また、教師無し学習即ち独習の状態では学習が飽和しているときには、ある期間だけでも教師について学習者の誤りが修正されることは効果的であることが確認された。なお、本研究は学習モデル自体をさらに追求していく一方、全体を実践的に進めていかなければならないと考える。

参考文献

- (1) 西, 降矢: “連想記憶モデルを用いた学習法とその効果の一考察”, 山口大学教育学部附属教育実践研究指導センタ研究紀要第1号(平元).
- (2) RHINE ユーザーズマニュアル.
- (3) 船橋賢一: “ニューラルネットワークのcapabilityについて”, 電子情報通信学会技術報告, MBE 88-52(1988).